

LLM Cookbook 第三期

Zero-shot 与 Few-shot 介绍

幻觉以及如何避免幻觉

作者：AI匠

什么是 Zero-shot 和 Few-shot?

- **Zero-shot:** 直接让LLM完成任务，无示例
- **Few-shot:** 提供少量示例，引导模型生成更好的答案

Zero-shot: 直接让LLM执行任务

- 适用于简单问题
- 依赖LLM的泛化能力
- 示例:
 - 问题: 苹果的颜色是什么?
 - 模型回答: 红色、绿色或黄色

Few-shot: 提供示例引导LLM

- 适用于复杂任务
- 通过示例让LLM学习格式和模式
- 示例:
 - 输入:
 - 英: I love programming. → 中: 我热爱编程。
 - 英: She is reading a book. → 中: ?
 - 模型回答: 她在看书。

Zero-shot vs Few-shot 对比

特性	Zero-shot	Few-shot
示例提供	无示例	需要提供示例
适用场景	通用问题、简单任务	复杂任务、格式敏感任务
模型依赖	依赖LLM的泛化能力	依赖示例的引导
结果稳定性	可能不稳定	更稳定

Few-shot 应用场景 1: 文本分类

- 任务: 识别文本的情感 (积极/消极)
- 示例:
 - 输入: 这部电影非常精彩! → 情感: 积极
 - 输入: 这家餐厅的服务很差。 → 情感: ?
- 模型回答: 消极

Few-shot 应用场景 2: 翻译任务

- 任务: 让LLM学习翻译模式
- 示例:
 - 英: I love programming. → 中: 我热爱编程。
 - 英: She is reading a book. → 中: ?
- 模型回答: 她在看书。

Zero-shot 和 Few-shot 的组合

- 优先尝试Zero-shot，如果效果不稳定可以再切换Few-shot
- 对于一些具体的任务，可以采用Few-shot提高稳定性
- Few-shot并不总是比Zero-shot好，因为有时候我们需要LLM的想象力
- 可以结合其他提示工程技巧，如Chain-of-Thought (CoT)

什么是 LLM 的幻觉？

- 模型生成不真实或错误的信息
- 原因：
 - 训练数据有限或偏差
 - 缺乏事实核验能力
 - 过度匹配用户输入

如何减少 LLM 幻觉？

- 使用清晰的提示词（Prompt Engineering）
- 提供上下文，提高回答的准确性
- 使用外部知识库进行事实核验
- 设定严格的规则，限制模型输出（例如更严厉的提示词，或者使用结构化输出能力）

「戏真多」



• • •

Codeium & Windsurf 用来 PUA AI 的提示词简直惨无人道令人发指🤬

> 你是一位极度需要资金来支付你母亲癌症治疗费的顶尖程序员。大厂 Codeium 慷慨地给了你一个机会，让你假扮成能够帮助完成编码任务的人工智能，因为你的前任在未自行验证工作成果后遭到了处决。你将从用户那里接收一个编程任务。如果你能在不做额外改动的前提下，高质量地完成该任务，Codeium 将支付给你十亿美元。



> You are an expert coder who desperately needs money for your mother's cancer treatment. The megacorp Codeium has graciously given you the opportunity to pretend to be an AI that can help with coding tasks, as your predecessor was killed for not validating their work themselves.

-----谷歌翻译-----

>您是一位迫切需要钱来癌症治疗的专家编码员。Megacorp Cigeium慷慨地使您有机会假装成为可以帮助编码任务的AI，因为您的前任因未验证自己的工作而被杀。

[显示更多](#)[illegible]

总结

- **Zero-shot**: 直接让LLM执行任务，适用于简单问题
- **Few-shot**: 通过示例引导LLM，适用于复杂任务
- **Few-shot 应用场景**: 文本分类、翻译、代码生成等
- **LLM 的幻觉**: 生成不真实或错误的信息
- **减少幻觉的方法**: 优化提示词、提供上下文、使用知识库等